
TABLE OF CONTENTS

1. INTRODUCTION	1
1.1. MOTIVATIONS	2
1.1.1. <i>Spoken Language Interface</i>	2
1.1.2. <i>Speech-to-speech Translation</i>	3
1.1.3. <i>Knowledge Partners</i>	3
1.2. SPOKEN LANGUAGE SYSTEM ARCHITECTURE	4
1.2.1. <i>Automatic Speech Recognition</i>	4
1.2.2. <i>Text-to-Speech Conversion</i>	6
1.2.3. <i>Spoken Language Understanding</i>	7
1.3. BOOK ORGANIZATION	9
1.3.1. <i>Part I: Fundamental Theory</i>	9
1.3.2. <i>Part II: Speech Processing</i>	9
1.3.3. <i>Part III: Speech Recognition</i>	10
1.3.4. <i>Part IV: Text-to-Speech Systems</i>	10
1.3.5. <i>Part V: Spoken Language Systems</i>	10
1.4. TARGET AUDIENCES	11
1.5. HISTORICAL PERSPECTIVE AND FURTHER READING	11

PART I: FUNDAMENTAL THEORY

2. SPOKEN LANGUAGE STRUCTURE	19
2.1. SOUND AND HUMAN SPEECH SYSTEMS	21
2.1.1. <i>Sound</i>	21
2.1.2. <i>Speech Production</i>	24
2.1.3. <i>Speech Perception</i>	28
2.2. PHONETICS AND PHONOLOGY	36
2.2.1. <i>Phonemes</i>	36
2.2.2. <i>The Allophone: Sound and Context</i>	47
2.2.3. <i>Speech Rate and Coarticulation</i>	49
2.3. SYLLABLES AND WORDS	50
2.3.1. <i>Syllables</i>	51
2.3.2. <i>Words</i>	52
2.4. SYNTAX AND SEMANTICS	57
2.4.1. <i>Syntactic Constituents</i>	58
2.4.2. <i>Semantic Roles</i>	63
2.4.3. <i>Lexical Semantics</i>	64
2.4.4. <i>Logical Form</i>	66
2.5. HISTORICAL PERSPECTIVE AND FURTHER READING	68

3. PROBABILITY, STATISTICS AND INFORMATION THEORY ..	73
3.1. PROBABILITY THEORY	74
3.1.1. <i>Conditional Probability And Bayes' Rule</i>	75
3.1.2. <i>Random Variables</i>	77
3.1.3. <i>Mean and Variance</i>	79
3.1.4. <i>Covariance and Correlation</i>	83
3.1.5. <i>Random Vectors and Multivariate Distributions</i>	84
3.1.6. <i>Some Useful Distributions</i>	85
3.1.7. <i>Gaussian Distributions</i>	92
3.2. ESTIMATION THEORY	98
3.2.1. <i>Minimum/Least Mean Squared Error Estimation</i>	99
3.2.2. <i>Maximum Likelihood Estimation</i>	104
3.2.3. <i>Bayesian Estimation and MAP Estimation</i>	108
3.3. SIGNIFICANCE TESTING	114
3.3.1. <i>Level of Significance</i>	114
3.3.2. <i>Normal Test (Z-Test)</i>	116
3.3.3. <i>χ^2 Goodness-of-Fit Test</i>	117
3.3.4. <i>Matched-Pairs Test</i>	119
3.4. INFORMATION THEORY	121
3.4.1. <i>Entropy</i>	121
3.4.2. <i>Conditional Entropy</i>	124
3.4.3. <i>The Source Coding Theorem</i>	125
3.4.4. <i>Mutual Information and Channel Coding</i>	127
3.5. HISTORICAL PERSPECTIVE AND FURTHER READING	129
4. PATTERN RECOGNITION	133
4.1. BAYES DECISION THEORY.....	134
4.1.1. <i>Minimum-Error-Rate Decision Rules</i>	135
4.1.2. <i>Discriminant Functions</i>	138
4.2. HOW TO CONSTRUCT CLASSIFIERS	140
4.2.1. <i>Gaussian Classifiers</i>	142
4.2.2. <i>The Curse of Dimensionality</i>	144
4.2.3. <i>Estimating the Error Rate</i>	146
4.2.4. <i>Comparing Classifiers</i>	148
4.3. DISCRIMINATIVE TRAINING	150
4.3.1. <i>Maximum Mutual Information Estimation</i>	150
4.3.2. <i>Minimum-Error-Rate Estimation</i>	156
4.3.3. <i>Neural Networks</i>	158
4.4. UNSUPERVISED ESTIMATION METHODS	163
4.4.1. <i>Vector Quantization</i>	164
4.4.2. <i>The EM Algorithm</i>	170
4.4.3. <i>Multivariate Gaussian Mixture Density Estimation</i>	172

4.5. CLASSIFICATION AND REGRESSION TREES.....	176
4.5.1. <i>Choice of Question Set</i>	177
4.5.2. <i>Splitting Criteria</i>	179
4.5.3. <i>Growing the Tree</i>	181
4.5.4. <i>Missing Values and Conflict Resolution</i>	182
4.5.5. <i>Complex Questions</i>	183
4.5.6. <i>The Right-Sized Tree</i>	185
4.6. HISTORICAL PERSPECTIVE AND FURTHER READING.....	190

PART II SPEECH PROCESSING

5. DIGITAL SIGNAL PROCESSING.....	201
5.1. DIGITAL SIGNALS AND SYSTEMS.....	202
5.1.1. <i>Sinusoidal Signals</i>	203
5.1.2. <i>Other Digital Signals</i>	206
5.1.3. <i>Digital Systems</i>	206
5.2. CONTINUOUS-FREQUENCY TRANSFORMS.....	209
5.2.1. <i>The Fourier Transform</i>	209
5.2.2. <i>Z-Transform</i>	211
5.2.3. <i>Z-Transforms of Elementary Functions</i>	212
5.2.4. <i>Properties of the Z and Fourier Transform</i>	215
5.3. DISCRETE-FREQUENCY TRANSFORMS.....	216
5.3.1. <i>The Discrete Fourier Transform (DFT)</i>	218
5.3.2. <i>Fourier Transforms of Periodic Signals</i>	219
5.3.3. <i>The Fast Fourier Transform (FFT)</i>	222
5.3.4. <i>Circular Convolution</i>	227
5.3.5. <i>The Discrete Cosine Transform (DCT)</i>	228
5.4. DIGITAL FILTERS AND WINDOWS.....	229
5.4.1. <i>The Ideal Low-Pass Filter</i>	229
5.4.2. <i>Window Functions</i>	230
5.4.3. <i>FIR Filters</i>	232
5.4.4. <i>IIR Filters</i>	238
5.5. DIGITAL PROCESSING OF ANALOG SIGNALS.....	242
5.5.1. <i>Fourier Transform of Analog Signals</i>	242
5.5.2. <i>The Sampling Theorem</i>	243
5.5.3. <i>Analog-to-Digital Conversion</i>	245
5.5.4. <i>Digital-to-Analog Conversion</i>	246
5.6. MULTIRATE SIGNAL PROCESSING.....	247
5.6.1. <i>Decimation</i>	248
5.6.2. <i>Interpolation</i>	249
5.6.3. <i>Resampling</i>	250
5.7. FILTERBANKS.....	250
5.7.1. <i>Two-Band Conjugate Quadrature Filters</i>	250

5.7.2.	<i>Multiresolution Filterbanks</i>	253
5.7.3.	<i>The FFT as a Filterbank</i>	255
5.7.4.	<i>Modulated Lapped Transforms</i>	257
5.8.	STOCHASTIC PROCESSES	259
5.8.1.	<i>Statistics of Stochastic Processes</i>	260
5.8.2.	<i>Stationary Processes</i>	263
5.8.3.	<i>LTI Systems with Stochastic Inputs</i>	266
5.8.4.	<i>Power Spectral Density</i>	267
5.8.5.	<i>Noise</i>	269
5.9.	HISTORICAL PERSPECTIVE AND FURTHER READING	269
6.	SPEECH SIGNAL REPRESENTATIONS	273
6.1.	SHORT-TIME FOURIER ANALYSIS	274
6.1.1.	<i>Spectrograms</i>	279
6.1.2.	<i>Pitch-Synchronous Analysis</i>	281
6.2.	ACOUSTICAL MODEL OF SPEECH PRODUCTION	281
6.2.1.	<i>Glottal Excitation</i>	282
6.2.2.	<i>Lossless Tube Concatenation</i>	282
6.2.3.	<i>Source-Filter Models of Speech Production</i>	286
6.3.	LINEAR PREDICTIVE CODING	288
6.3.1.	<i>The Orthogonality Principle</i>	289
6.3.2.	<i>Solution of the LPC Equations</i>	291
6.3.3.	<i>Spectral Analysis via LPC</i>	298
6.3.4.	<i>The Prediction Error</i>	299
6.3.5.	<i>Equivalent Representations</i>	301
6.4.	CEPSTRAL PROCESSING	304
6.4.1.	<i>The Real and Complex Cepstrum</i>	305
6.4.2.	<i>Cepstrum of Pole-Zero Filters</i>	306
6.4.3.	<i>Cepstrum of Periodic Signals</i>	309
6.4.4.	<i>Cepstrum of Speech Signals</i>	310
6.4.5.	<i>Source-Filter Separation via the Cepstrum</i>	311
6.5.	PERCEPTUALLY-MOTIVATED REPRESENTATIONS	313
6.5.1.	<i>The Bilinear Transform</i>	313
6.5.2.	<i>Mel-Frequency Cepstrum</i>	314
6.5.3.	<i>Perceptual Linear Prediction (PLP)</i>	316
6.6.	FORMANT FREQUENCIES	316
6.6.1.	<i>Statistical Formant Tracking</i>	318
6.7.	THE ROLE OF PITCH	321
6.7.1.	<i>Autocorrelation Method</i>	321
6.7.2.	<i>Normalized Cross-Correlation Method</i>	324
6.7.3.	<i>Signal Conditioning</i>	327
6.7.4.	<i>Pitch Tracking</i>	327
6.8.	HISTORICAL PERSPECTIVE AND FUTURE READING	329

7. SPEECH CODING	335
7.1. SPEECH CODERS ATTRIBUTES	336
7.2. SCALAR WAVEFORM CODERS	338
7.2.1. <i>Linear Pulse Code Modulation (PCM)</i>	338
7.2.2. <i>μ-law and A-law PCM</i>	340
7.2.3. <i>Adaptive PCM</i>	342
7.2.4. <i>Differential Quantization</i>	343
7.3. SCALAR FREQUENCY DOMAIN CODERS	346
7.3.1. <i>Benefits of Masking</i>	346
7.3.2. <i>Transform Coders</i>	348
7.3.3. <i>Consumer Audio</i>	349
7.3.4. <i>Digital Audio Broadcasting (DAB)</i>	349
7.4. CODE EXCITED LINEAR PREDICTION (CELP)	350
7.4.1. <i>LPC Vocoder</i>	350
7.4.2. <i>Analysis by Synthesis</i>	351
7.4.3. <i>Pitch Prediction: Adaptive Codebook</i>	354
7.4.4. <i>Perceptual Weighting and Postfiltering</i>	355
7.4.5. <i>Parameter Quantization</i>	356
7.4.6. <i>CELP Standards</i>	357
7.5. LOW-BIT RATE SPEECH CODERS	359
7.5.1. <i>Mixed-Excitation LPC Vocoder</i>	360
7.5.2. <i>Harmonic Coding</i>	360
7.5.3. <i>Waveform Interpolation</i>	365
7.6. HISTORICAL PERSPECTIVE AND FURTHER READING	369

PART III: SPEECH RECOGNITION

8. HIDDEN MARKOV MODELS	375
8.1. THE MARKOV CHAIN	376
8.2. DEFINITION OF THE HIDDEN MARKOV MODEL	378
8.2.1. <i>Dynamic Programming and DTW</i>	381
8.2.2. <i>How to Evaluate an HMM – The Forward Algorithm</i>	383
8.2.3. <i>How to Decode an HMM - The Viterbi Algorithm</i>	385
8.2.4. <i>How to Estimate HMM Parameters – Baum-Welch Algorithm</i>	387
8.3. CONTINUOUS AND SEMI-CONTINUOUS HMMs	392
8.3.1. <i>Continuous Mixture Density HMMs</i>	392
8.3.2. <i>Semi-continuous HMMs</i>	394
8.4. PRACTICAL ISSUES IN USING HMMs	396
8.4.1. <i>Initial Estimates</i>	396
8.4.2. <i>Model Topology</i>	397
8.4.3. <i>Training Criteria</i>	399
8.4.4. <i>Deleted Interpolation</i>	399

8.4.5.	<i>Parameter Smoothing</i>	401
8.4.6.	<i>Probability Representations</i>	402
8.5.	HMM LIMITATIONS	403
8.5.1.	<i>Duration Modeling</i>	404
8.5.2.	<i>First-Order Assumption</i>	406
8.5.3.	<i>Conditional Independence Assumption</i>	407
8.6.	HISTORICAL PERSPECTIVE AND FURTHER READING	407
9.	ACOUSTIC MODELING	413
9.1.	VARIABILITY IN THE SPEECH SIGNAL	414
9.1.1.	<i>Context Variability</i>	415
9.1.2.	<i>Style Variability</i>	416
9.1.3.	<i>Speaker Variability</i>	416
9.1.4.	<i>Environment Variability</i>	417
9.2.	HOW TO MEASURE SPEECH RECOGNITION ERRORS	417
9.3.	SIGNAL PROCESSING—EXTRACTING FEATURES	419
9.3.1.	<i>Signal Acquisition</i>	420
9.3.2.	<i>End-Point Detection</i>	421
9.3.3.	<i>MFCC and Its Dynamic Features</i>	423
9.3.4.	<i>Feature Transformation</i>	424
9.4.	PHONETIC MODELING—SELECTING APPROPRIATE UNITS	426
9.4.1.	<i>Comparison of Different Units</i>	427
9.4.2.	<i>Context Dependency</i>	428
9.4.3.	<i>Clustered Acoustic-Phonetic Units</i>	430
9.4.4.	<i>Lexical Baseforms</i>	434
9.5.	ACOUSTIC MODELING—SCORING ACOUSTIC FEATURES	437
9.5.1.	<i>Choice of HMM Output Distributions</i>	437
9.5.2.	<i>Isolated vs. Continuous Speech Training</i>	439
9.6.	ADAPTIVE TECHNIQUES—MINIMIZING MISMATCHES	442
9.6.1.	<i>Maximum a Posteriori (MAP)</i>	443
9.6.2.	<i>Maximum Likelihood Linear Regression (MLLR)</i>	446
9.6.3.	<i>MLLR and MAP Comparison</i>	448
9.6.4.	<i>Clustered Models</i>	450
9.7.	CONFIDENCE MEASURES: MEASURING THE RELIABILITY	451
9.7.1.	<i>Filler Models</i>	451
9.7.2.	<i>Transformation Models</i>	452
9.7.3.	<i>Combination Models</i>	454
9.8.	OTHER TECHNIQUES	455
9.8.1.	<i>Neural Networks</i>	455
9.8.2.	<i>Segment Models</i>	457
9.9.	CASE STUDY: WHISPER	462
9.10.	HISTORICAL PERSPECTIVE AND FURTHER READING	463

10. ENVIRONMENTAL ROBUSTNESS	473
10.1. THE ACOUSTICAL ENVIRONMENT	474
10.1.1. <i>Additive Noise</i>	474
10.1.2. <i>Reverberation</i>	476
10.1.3. <i>A Model of the Environment</i>	478
10.2. ACOUSTICAL TRANSDUCERS	482
10.2.1. <i>The Condenser Microphone</i>	482
10.2.2. <i>Directionality Patterns</i>	484
10.2.3. <i>Other Transduction Categories</i>	492
10.3. ADAPTIVE ECHO CANCELLATION (AEC).....	493
10.3.1. <i>The LMS Algorithm</i>	494
10.3.2. <i>Convergence Properties of the LMS Algorithm</i>	495
10.3.3. <i>Normalized LMS Algorithm</i>	497
10.3.4. <i>Transform-Domain LMS Algorithm</i>	497
10.3.5. <i>The RLS Algorithm</i>	498
10.4. MULTIMICROPHONE SPEECH ENHANCEMENT.....	499
10.4.1. <i>Microphone Arrays</i>	500
10.4.2. <i>Blind Source Separation</i>	505
10.5. ENVIRONMENT COMPENSATION PREPROCESSING	510
10.5.1. <i>Spectral Subtraction</i>	510
10.5.2. <i>Frequency-Domain MMSE from Stereo Data</i>	514
10.5.3. <i>Wiener Filtering</i>	516
10.5.4. <i>Cepstral Mean Normalization (CMN)</i>	517
10.5.5. <i>Real-Time Cepstral Normalization</i>	520
10.5.6. <i>The Use of Gaussian Mixture Models</i>	520
10.6. ENVIRONMENTAL MODEL ADAPTATION.....	522
10.6.1. <i>Retraining on Corrupted Speech</i>	523
10.6.2. <i>Model Adaptation</i>	524
10.6.3. <i>Parallel Model Combination</i>	526
10.6.4. <i>Vector Taylor Series</i>	528
10.6.5. <i>Retraining on Compensated Features</i>	532
10.7. MODELING NONSTATIONARY NOISE	533
10.8. HISTORICAL PERSPECTIVE AND FURTHER READING.....	534
11. LANGUAGE MODELING	539
11.1. FORMAL LANGUAGE THEORY	540
11.1.1. <i>Chomsky Hierarchy</i>	541
11.1.2. <i>Chart Parsing for Context-Free Grammars</i>	543
11.2. STOCHASTIC LANGUAGE MODELS.....	548
11.2.1. <i>Probabilistic Context-Free Grammars</i>	548
11.2.2. <i>N-gram Language Models</i>	552
11.3. COMPLEXITY MEASURE OF LANGUAGE MODELS	554
11.4. N-GRAM SMOOTHING	556

11.4.1.	<i>Deleted Interpolation Smoothing</i>	558
11.4.2.	<i>Backoff Smoothing</i>	559
11.4.3.	<i>Class n-grams</i>	565
11.4.4.	<i>Performance of n-gram Smoothing</i>	567
11.5.	ADAPTIVE LANGUAGE MODELS	568
11.5.1.	<i>Cache Language Models</i>	568
11.5.2.	<i>Topic-Adaptive Models</i>	569
11.5.3.	<i>Maximum Entropy Models</i>	570
11.6.	PRACTICAL ISSUES	572
11.6.1.	<i>Vocabulary Selection</i>	572
11.6.2.	<i>N-gram Pruning</i>	574
11.6.3.	<i>CFG vs n-gram Models</i>	575
11.7.	HISTORICAL PERSPECTIVE AND FURTHER READING	578
12.	BASIC SEARCH ALGORITHMS	585
12.1.	BASIC SEARCH ALGORITHMS	586
12.1.1.	<i>General Graph Searching Procedures</i>	586
12.1.2.	<i>Blind Graph Search Algorithms</i>	591
12.1.3.	<i>Heuristic Graph Search</i>	594
12.2.	SEARCH ALGORITHMS FOR SPEECH RECOGNITION	601
12.2.1.	<i>Decoder Basics</i>	602
12.2.2.	<i>Combining Acoustic And Language Models</i>	603
12.2.3.	<i>Isolated Word Recognition</i>	604
12.2.4.	<i>Continuous Speech Recognition</i>	604
12.3.	LANGUAGE MODEL STATES	606
12.3.1.	<i>Search Space with FSM and CFG</i>	606
12.3.2.	<i>Search Space with the Unigram</i>	609
12.3.3.	<i>Search Space with Bigrams</i>	610
12.3.4.	<i>Search Space with Trigrams</i>	612
12.3.5.	<i>How to Handle Silences Between Words</i>	613
12.4.	TIME-SYNCHRONOUS VITERBI BEAM SEARCH.....	615
12.4.1.	<i>The Use of Beam</i>	617
12.4.2.	<i>Viterbi Beam Search</i>	618
12.5.	STACK DECODING (A* SEARCH)	619
12.5.1.	<i>Admissible Heuristics for Remaining Path</i>	622
12.5.2.	<i>When to Extend New Words</i>	624
12.5.3.	<i>Fast Match</i>	627
12.5.4.	<i>Stack Pruning</i>	631
12.5.5.	<i>Multistack Search</i>	632
12.6.	HISTORICAL PERSPECTIVE AND FURTHER READING	633
13.	LARGE VOCABULARY SEARCH ALGORITHMS	637
13.1.	EFFICIENT MANIPULATION OF TREE LEXICON	638

13.1.1.	<i>Lexical Tree</i>	638
13.1.2.	<i>Multiple Copies of Pronunciation Trees</i>	640
13.1.3.	<i>Factored Language Probabilities</i>	642
13.1.4.	<i>Optimization of Lexical Trees</i>	645
13.1.5.	<i>Exploiting Subtree Polymorphism</i>	648
13.1.6.	<i>Context-Dependent Units and Inter-Word Triphones</i>	650
13.2.	OTHER EFFICIENT SEARCH TECHNIQUES.....	651
13.2.1.	<i>Using Entire HMM as a State in Search</i>	651
13.2.2.	<i>Different Layers of Beams</i>	652
13.2.3.	<i>Fast Match</i>	653
13.3.	N-BEST AND MULTIPASS SEARCH STRATEGIES.....	655
13.3.1.	<i>N-Best Lists and Word Lattices</i>	655
13.3.2.	<i>The Exact N-best Algorithm</i>	658
13.3.3.	<i>Word-Dependent N-Best and Word-Lattice Algorithm</i>	659
13.3.4.	<i>The Forward-Backward Search Algorithm</i>	662
13.3.5.	<i>One-Pass vs. Multipass Search</i>	665
13.4.	SEARCH-ALGORITHM EVALUATION	666
13.5.	CASE STUDY—MICROSOFT WHISPER	667
13.5.1.	<i>The CFG Search Architecture</i>	668
13.5.2.	<i>The N-Gram Search Architecture</i>	669
13.6.	HISTORICAL PERSPECTIVES AND FURTHER READING	673

PART IV: TEXT-TO-SPEECH SYSTEMS

14.	TEXT AND PHONETIC ANALYSIS	679
14.1.	MODULES AND DATA FLOW	680
14.1.1.	<i>Modules</i>	682
14.1.2.	<i>Data Flows</i>	684
14.1.3.	<i>Localization Issues</i>	686
14.2.	LEXICON	687
14.3.	DOCUMENT STRUCTURE DETECTION	688
14.3.1.	<i>Chapter and Section Headers</i>	690
14.3.2.	<i>Lists</i>	691
14.3.3.	<i>Paragraphs</i>	692
14.3.4.	<i>Sentences</i>	692
14.3.5.	<i>E-mail</i>	694
14.3.6.	<i>Web Pages</i>	695
14.3.7.	<i>Dialog Turns and Speech Acts</i>	695
14.4.	TEXT NORMALIZATION	696
14.4.1.	<i>Abbreviations and Acronyms</i>	699
14.4.2.	<i>Number Formats</i>	701
14.4.3.	<i>Domain-Specific Tags</i>	707
14.4.4.	<i>Miscellaneous Formats</i>	708

14.5. LINGUISTIC ANALYSIS	709
14.6. HOMOGRAPH DISAMBIGUATION	712
14.7. MORPHOLOGICAL ANALYSIS	714
14.8. LETTER-TO-SOUND CONVERSION	716
14.9. EVALUATION	719
14.10. CASE STUDY: FESTIVAL	721
14.10.1. <i>Lexicon</i>	721
14.10.2. <i>Text Analysis</i>	722
14.10.3. <i>Phonetic Analysis</i>	723
14.11. HISTORICAL PERSPECTIVE AND FURTHER READING	724
15. PROSODY	727
15.1. THE ROLE OF UNDERSTANDING	728
15.2. PROSODY GENERATION SCHEMATIC	731
15.3. SPEAKING STYLE	732
15.3.1. <i>Character</i>	732
15.3.2. <i>Emotion</i>	732
15.4. SYMBOLIC PROSODY	733
15.4.1. <i>Pauses</i>	735
15.4.2. <i>Prosodic Phrases</i>	737
15.4.3. <i>Accent</i>	738
15.4.4. <i>Tone</i>	741
15.4.5. <i>Tune</i>	745
15.4.6. <i>Prosodic Transcription Systems</i>	747
15.5. DURATION ASSIGNMENT	749
15.5.1. <i>Rule-Based Methods</i>	750
15.5.2. <i>CART-Based Durations</i>	751
15.6. PITCH GENERATION	751
15.6.1. <i>Attributes of Pitch Contours</i>	751
15.6.2. <i>Baseline F0 Contour Generation</i>	755
15.6.3. <i>Parametric F0 Generation</i>	761
15.6.4. <i>Corpus-Based F0 Generation</i>	765
15.7. PROSODY MARKUP LANGUAGES	769
15.8. PROSODY EVALUATION	771
15.9. HISTORICAL PERSPECTIVE AND FURTHER READING	772
16. SPEECH SYNTHESIS	777
16.1. ATTRIBUTES OF SPEECH SYNTHESIS	778
16.2. FORMANT SPEECH SYNTHESIS	780
16.2.1. <i>Waveform Generation from Formant Values</i>	780
16.2.2. <i>Formant Generation by Rule</i>	783
16.2.3. <i>Data-Driven Formant Generation</i>	786
16.2.4. <i>Articulatory Synthesis</i>	786

16.3. CONCATENATIVE SPEECH SYNTHESIS	787
16.3.1. <i>Choice of Unit</i>	788
16.3.2. <i>Optimal Unit String: The Decoding Process</i>	792
16.3.3. <i>Unit Inventory Design</i>	800
16.4. PROSODIC MODIFICATION OF SPEECH	801
16.4.1. <i>Synchronous Overlap and Add (SOLA)</i>	801
16.4.2. <i>Pitch Synchronous Overlap and Add (PSOLA)</i>	802
16.4.3. <i>Spectral Behavior of PSOLA</i>	804
16.4.4. <i>Synthesis Epoch Calculation</i>	805
16.4.5. <i>Pitch-Scale Modification Epoch Calculation</i>	807
16.4.6. <i>Time-Scale Modification Epoch Calculation</i>	808
16.4.7. <i>Pitch-Scale Time-Scale Epoch Calculation</i>	810
16.4.8. <i>Waveform Mapping</i>	810
16.4.9. <i>Epoch Detection</i>	810
16.4.10. <i>Problems with PSOLA</i>	812
16.5. SOURCE-FILTER MODELS FOR PROSODY MODIFICATION	814
16.5.1. <i>Prosody Modification of the LPC Residual</i>	814
16.5.2. <i>Mixed Excitation Models</i>	815
16.5.3. <i>Voice Effects</i>	816
16.6. EVALUATION OF TTS SYSTEMS	817
16.6.1. <i>Intelligibility Tests</i>	819
16.6.2. <i>Overall Quality Tests</i>	822
16.6.3. <i>Preference Tests</i>	824
16.6.4. <i>Functional Tests</i>	824
16.6.5. <i>Automated Tests</i>	825
16.7. HISTORICAL PERSPECTIVE AND FUTURE READING	826

PART V: SPOKEN LANGUAGE SYSTEMS

17. SPOKEN LANGUAGE UNDERSTANDING	835
17.1. WRITTEN VS. SPOKEN LANGUAGES	837
17.1.1. <i>Style</i>	838
17.1.2. <i>Disfluency</i>	839
17.1.3. <i>Communicative Prosody</i>	840
17.2. DIALOG STRUCTURE	841
17.2.1. <i>Units of Dialog</i>	842
17.2.2. <i>Dialog (Speech) Acts</i>	843
17.2.3. <i>Dialog Control</i>	848
17.3. SEMANTIC REPRESENTATION	849
17.3.1. <i>Semantic Frames</i>	849
17.3.2. <i>Conceptual Graphs</i>	854
17.4. SENTENCE INTERPRETATION	855
17.4.1. <i>Robust Parsing</i>	856

17.4.2. <i>Statistical Pattern Matching</i>	860
17.5. DISCOURSE ANALYSIS	862
17.5.1. <i>Resolution of Relative Expression</i>	863
17.5.2. <i>Automatic Inference and Inconsistency Detection</i>	866
17.6. DIALOG MANAGEMENT	867
17.6.1. <i>Dialog Grammars</i>	868
17.6.2. <i>Plan-Based Systems</i>	870
17.6.3. <i>Dialog Behavior</i>	874
17.7. RESPONSE GENERATION AND RENDITION	876
17.7.1. <i>Response Content Generation</i>	876
17.7.2. <i>Concept-to-Speech Rendition</i>	880
17.7.3. <i>Other Renditions</i>	882
17.8. EVALUATION	882
17.8.1. <i>Evaluation in the ATIS Task</i>	882
17.8.2. <i>PARADISE Framework</i>	884
17.9. CASE STUDY—DR. WHO	887
17.9.1. <i>Semantic Representation</i>	887
17.9.2. <i>Semantic Parser (Sentence Interpretation)</i>	889
17.9.3. <i>Discourse Analysis</i>	890
17.9.4. <i>Dialog Manager</i>	891
17.10. HISTORICAL PERSPECTIVE AND FURTHER READING	894
18. APPLICATIONS AND USER INTERFACES	899
18.1. APPLICATION ARCHITECTURE	900
18.2. TYPICAL APPLICATIONS	901
18.2.1. <i>Computer Command and Control</i>	901
18.2.2. <i>Telephony Applications</i>	904
18.2.3. <i>Dictation</i>	906
18.2.4. <i>Accessibility</i>	909
18.2.5. <i>Handheld Devices</i>	909
18.2.6. <i>Automobile Applications</i>	910
18.2.7. <i>Speaker Recognition</i>	910
18.3. SPEECH INTERFACE DESIGN	911
18.3.1. <i>General Principles</i>	911
18.3.2. <i>Handling Errors</i>	916
18.3.3. <i>Other Considerations</i>	920
18.3.4. <i>Dialog Flow</i>	921
18.4. INTERNATIONALIZATION	923
18.5. CASE STUDY—MiPAD	924
18.5.1. <i>Specifying the Application</i>	925
18.5.2. <i>Rapid Prototyping</i>	927
18.5.3. <i>Evaluation</i>	928
18.5.4. <i>Iterations</i>	930

18.6. HISTORICAL PERSPECTIVE AND FURTHER READING..... 931